CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	Multimodal sensor fusion for low-power wearable human motion tracking systems in sports applications
Authors	Wilk, Mariusz P.;Walsh, Michael;O'Flynn, Brendan
Publication date	2020-10-13
Original Citation	Wilk, M. P., Walsh, M. and O'Flynn, B. (2020) 'Multimodal Sensor Fusion for Low-Power Wearable Human Motion Tracking Systems in Sports Applications', IEEE Sensors Journal, doi: 10.1109/JSEN.2020.3030779
Type of publication	Article (peer-reviewed)
Link to publisher's version	https://ieeexplore.ieee.org/document/9222148 - 10.1109/JSEN.2020.3030779
Rights	© 2020 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works
Download date	2025-07-01 21:17:58
Item downloaded from	https://hdl.handle.net/10468/10666

UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

Multimodal Sensor Fusion for Low-Power Wearable Human Motion Tracking Systems in Sports Applications

Mariusz P. Wilk, *Member, IEEE,* Michael Walsh, *Member, IEEE,* Brendan O'Flynn, *Member, IEEE*

***Abstract*— This paper presents a prototype human motion tracking system for wearable sports applications. It can be particularly applicable for tracking human motion during executing certain strength training exercises, such as the barbell squat, where an inappropriate technique could result in an injury. The key novelty of the proposed system is twofold. Firstly, it is an inside-out, multimodal, motion tracker that incorporates two complementary sensor modalities, i.e. a camera and an inertial motion sensor, as well as two externally-mounted points of reference. Secondly, it incorporates a novel multimodal sensor fusion algorithm which uses the complementary nature of vision and inertial sensor modalities to perform a computationally efficient 3-Dimensional (3-D) pose detection of the wearable device. The 3-D pose is determined by fusing information about the two external reference points captured by the camera together with the orientation angles captured by the inertial motion sensor. The accuracy of the prototype was experimentally validated in laboratory conditions. The main findings are as follows. The Root Mean Square Error (RMSE) in 3-D position calculation was 36.7 mm and 13.6 mm in the static and mobile cases, respectively. Whereas the static case was aimed at determining the system's performance at all 3-D poses within the work envelope, the mobile case was used to determine the error in tracking human motion that is involved in the barbell squat, i.e. a mainly repeated vertical motion pattern.**

***Index Terms*— Inertial Motion Sensor, Inside-Out Tracking, Pose Detection, Monocular Camera, Multimodal, Sensor Fusion, 3-D**

I. Introduction

MULTIMODAL sensor data fusion is a common approach to solving problems in applications wherein a single sensor modality fails to provide enough information to solve the given problem. In such cases, sensors with different complementary modalities are often used together to overcome this difficulty. The complementary nature of certain sensor modalities can be helpful for tackling problems that would be difficult to solve otherwise. One of the most common examples includes the Inertial Motion Unit (IMU) sensor, which is a multi-sensor, multimodal, device in a single package. It comprises three sensor modalities that complement each other's weaknesses, i.e. the accelerometer, magnetometer, and gyroscope. Despite that, it is often considered as a single device whose three sensor modalities are fused together to produce a reliable orientation measurement using an algorithm, such as that based on the Gradient Descent [1]. Another example of such a complementary set of sensor modalities is the combination of vision with the IMU sensors. The vision sensor technology can provide information that the IMU cannot capture and vice versa. For example, the camera can be used to determine the absolute position of a given point in space. It is difficult to achieve that with an IMU due to its inherent limitations, such as the drift or the disturbances in magnetic field. Likewise, the IMU can capture motion independently of the lighting conditions or occlusions, which are some of the main weaknesses of the vision sensors. Even the most advanced image processing algorithms may prove ineffective under adverse or unexpected lighting conditions. Although the vision sensors can be used to effectively track points of interest in their Field-of-View (FoV), they can quickly lose the

This publication has emanated from research conducted with the financial support of Science Foundation Ireland (SFI) and is co-funded under the European Regional Development Fund under Grant Number 13/RC/2077.

Mariusz P. Wilk is currently with Watertools Ltd, Cork, Ireland (email: mariusz@watertols.com)

Michael Walsh is with the Tyndall National Institute, University College Cork, Ireland (email: michael.walsh @tyndall.ie)

Brendan O'Flynn is with the Tyndall National Institute, University College Cork, Ireland (email: brendan.oflynn@tyndall.ie)

XXXX-XXXX © XXXX IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

tracking if the line-of-sight is not maintained for most of the time. The resulting intermittent tracking can give rise problems such as the correspondence problem, i.e. losing track of which point of interest is which. Though, solutions to this problem exist, such as that based on amplitude modulation of the LED-based active markers [2].

Human motion tracking is a term that describes the process of detecting and tracking motion of the human body. The 3-Dimensional (3-D) pose detection is one of the tasks in this process. It involves finding the position and orientation of an object in 3-D space; also referred to as the 6-Degree-Of-Freedom (6-DOF) pose. Human motion tracking can be considered a largely solved problem if it is carried out under controlled conditions with virtually unlimited processing resources. Some of the common examples of such systems include the marker-based VICON (passive reflective markers) or the Optotrack (active markers) systems. These are often considered a 'Gold Standard' with the position accuracy of approximately 1 mm [3, 4]. However, it is not such a straightforward task to reliably track the 3-D pose of the human body, or its parts, in the context of highly resource-constrained systems; such as those where power consumption or ease of use are an important consideration. Moreover, it is not feasible to use infrastructure-heavy system setups, such as those used in the VICON-like systems, in such application spaces or as part of in-field experiments.

An example of a low-power system with little to no infrastructure requirements would include a wearable motion tracker used certain sports applications, such as the Strength Training (ST). ST was shown to be an important addition to regular exercise routines that can offset the negative effects of our increasingly sedentary lifestyle. It can also help delay and ease many of the age-related problems, such as cognitive decline, osteoarthritis, sarcopenia, to name a few [5]. However, an ST routine must be executed correctly to be effective and safe. There is an inherent risk of injury associated with it; particularly among people with little prior experience in ST; especially the older people. Some of the most effective exercises can be dangerous if carried out incorrectly, such as the barbell squats. The barbell squat involves compound movements and often significant weights, which increases the risk and seriousness of a potential injury. For example, the lumbar spine can be at a high risk of injury if the forward trunk lean is too high while executing the squat [6, 7]. The risk of knee and hip injury significantly increase if the squats are too deep and/or the lateral hip shift occurs [8-11]. Therefore, the supervision of a professional coach is necessary. This can be a challenge as the accessibility and affordability of the coaches can be limited and expensive; with the growing proportion of older population. Technology can help to ease this challenge. It can be used for tracking the correctness of the execution of certain exercises and provide feedback in real time. A low-power and highly miniaturised wearable human motion tracking system can be useful in such applications. In terms of accuracy, such a system should be sufficiently accurate to reliably track the motion in the particular exercise. The specific quantitative requirements or recommendations, as to the permitted error level, are not found in the existing literature. It is so because the exercise assessments are generally carried out by the couches subjectively on an individual basis following general guidelines. However, an approximate requirement for error in position tracking in the barbell squat can be estimated, based on the ranges of motion involved in this exercise. Some of the key parameters used in ensuring that the squat is executed correctly can be used as the basis for forming this requirement. For example, the vertical range of motion in a squat carried out by an average adult individual may vary between approximately 0.5 m and 1 m, which is used in measuring the squat's depth. Likewise, the forward trunk lean can be measured by tracking the position and orientation of the line segment between two points on the back, i.e. a 3-D vector's endpoint is on the upper back, below the barbell, and origin in the lower back, on the sacral section of the spine. While the magnitude of this vector would not vary significantly, the values of its individual components would; especially those along the vertical and the forward-facing horizontal components of the 3-D position, i.e. the y and z, respectively. The distance between these two points on the adult athlete's back can be assumed to be approximately 0.5 m. The angle between this vector and the floor can vary between 45 degrees and 90 degrees [12]. Therefore, the values of this vector's components y and z would vary by up to approximately 35 cm. In the case of lateral hip shift, the range of motion would be smaller. It would normally reach up to a half the distance between the two feet, i.e. approximately 30 cm for an adult. Therefore, the error in position tracking of the motion tracker, referred to in this work as the Wearable Platform (WP), would be expected to remain at sub-centimetre level. However, not only should the motion tracking system have very low error in its measurements to meet these requirements, but it should also be affordable, simple to set up, and easy to use. Hence, the right balance between these factors is desired.

The main contribution of this work is the novel wearable motion tracking system that can meet the requirements of low power applications such as those in ST. The key novelty of the proposed system is twofold.

XXXX-XXXX © XXXX IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

Firstly, the WP is an opto-inertial tracker that can determine the 3-D pose using information from its two complementary sensor modalities, i.e. a monocular camera with an IMU, and only two externally mounted points of reference in the form of IR LEDs. The proposed WP is an inside-out tracker, i.e. the camera is embedded in the wearable unit itself to track the two external IR LEDs. The second main novelty of this work is the multimodal sensor fusion algorithm for computationally efficient 3-D motion tracking. The algorithm was designed such that it can be executed on wearable motion trackers with limited processing power. The proposed algorithm fuses the information about the two reference points, captured by the camera, together with the orientation angles from the IMU. The system architecture was designed with the algorithm in mind and vice versa. This paper describes the hardware-software co-design approach required to develop the inside-out 3-D motion tracking system. This work significantly advances the State-Of-the-Art (SOA) by proposing a novel approach to motion tracking that combines the advantages of the existing alternatives. It is unique in several ways. Like the leading IS-1500 system, it is an inside-out, opto-inertial, motion tracker [13]. However, it requires only two external points of reference to compute the 3-D pose, which reduces the computational complexity. The computational complexity reduction is achieved with the proposed data fusion algorithm which uses the geometric structures that are formed between the camera's principal point and the two reference points and complements the missing pieces of information with orientation angles from the IMU. In this sense, the proposed algorithm is similar to the approach of two outside-in trackers in the literature that also use two reference points (outside-in trackers use externally mounted cameras), i.e. those proposed by Maereg et al. and Li et al. [14, 15]. The accuracy in 3-D position tracking of the proposed system did not match that of the IS-1500, but it was comparable with the other two trackers. The Root Mean Square Error (RMSE) in 3-D position was 36.7 mm and 13.6 mm in the static and mobile cases, respectively. Therefore, the significance of this work is in that the proposed system architecture, along with the purpose-designed data fusion algorithm, offer a motion tracker that operates similarly to the IS-1500 but with lower overall requirements. The proposed approach can be considered an enabling factor towards performing motion tracking functions in cost-sensitive applications where a balance between tracking accuracy, processing speed, form factor, ease of use, and the cost is required.

The paper is organized as follows. The broader research context and the SOA analysis are described in Section I and Section II, respectively. The main body of this work is described in detail in Section III. It includes the proposed system architecture (Section III A), the proposed sensor fusion algorithm (Section III B) and the description of the simulations and experimental work, which include both static and mobile scenarios (Sections III C, and III D, respectively). The main experimental results in static and mobile scenario are shown in

TABLE III and TABLE IV, respectively. The discussion on the results and how they compared to similar systems found SOA is found in Section V. The work is concluded with a summary of the main findings and recommendations for future work (see Section VI).

II. Previous Work

Human motion tracking is a popular topic in the scientific community. A body of scientific literature shows the use of both unimodal and multimodal sensor approaches for this purpose. Unimodal approaches focus on using sensors with a single modality, e.g. cameras, or accelerometers. Some of the most common approaches include vision sensor technology. In recent years, the introduction of depth cameras sparked a revolution in motion tracking using low-cost consumer-grade devices, such as the Microsoft's Kinect™ V1 released in 2010 and the improved Kinect™ V2, released in 2013. It was embraced by the scientific community mainly due to its 3-D skeletal tracking capabilities [16]. It was shown to be a viable tool for biomechanical gait analysis. Its performance was compared to the earlier mentioned VICON motion capture system [17]. Another example of unimodal motion tracking that is worth noting, includes the use of wearable IMU sensors. Although IMUs are not strictly unimodal sensor systems, their output may be treated as such. In that context, the IMUs can be used effectively for motion tracking. One of the more notable examples is the Xsens MVN motion capture whole-body suit [18]. It is a marker-less and camera-less system that uses wearable IMUs with sensor fusion algorithms. IMUs are widely used in various motion tracking technologies, such as the data gloves [19]. There exist other sensor technologies that are used in motion tracking and positioning applications, such as those involving the acoustic, radio frequency, or time-of-flight techniques. However, the vision and IMU based techniques tend to dominate this application space.

Multimodal data fusion techniques are often preferred when a single sensor modality proves insufficient to obtain enough information to reliably perform the motion tracking function. It is often necessary in the context of highly miniaturized, low-power, wearable devices for

XXXX-XXXX © XXXX IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

motion tracking. The most common sensor modality choices in the literature include the combination of both vision and IMU sensors. These choices are motivated by their complementary nature, i.e. each sensor type provides data that complement the weaknesses of the other sensor type in the system, and vice-versa (i.e. drifts for IMUs and occlusions or lighting conditions for cameras in this case). The combination of these two sensor modalities, in conjunction with sensor fusion algorithms, can result in a reliable 6-DOF pose detection. One of the most notable advances in the SOA is the inclusion of a camera in the wearable device itself. Such motion tracking systems are generally referred to as inside-out trackers, which is one of two main categories of such systems, i.e. inside-out and outside-in. Whereas in the inside-out trackers the camera is attached to the tracked object, whose position and orientation in 3-D space is determined based on externally mounted points of reference, the outside-in systems use an externally mounted camera to track points of reference attached to the moving object of interest [20]. An example of both sensor modalities embedded in the wearable motion tracking devices for 6-DOF pose detection was proposed by Foxlin et al. [21-23], including their latest product IS-1500 [13]. These are the inside-out tracking systems that use a monocular camera (single camera) to track multiple fiducial markers embedded in the ambient environment and an IMU to correct for the motion and occlusions. The IS-1500 does have a number of limitations. It requires at least four reference points and has high processing requirements. Although the IS-1500 tracker has a miniature form factor, it requires an external processing unit with significant computational power, e.g. a laptop PC. Other examples include outside-in tracking systems where a monocular camera was embedded in the ambient environment to track two points of reference attached to a mobile/moving device that also incorporated an IMU; as described by Maereg et al. and Li et al. [14, 15]. These two outside-in systems have lower processing requirements, but they require an external camera with the two reference points having to be located on the object of interest. It makes these devices less practical in the considered application space.

These works show the evidence for an emerging trend in 3-D pose detection methods that increasingly incorporate monocular vision and IMU sensors in a single wearable unit. The wearable unit is effectively a wearable smart sensor that is driven by the multimodal sensor fusion algorithms. The advances in the SOA in camera miniaturization [24, 25] are accompanied by algorithms that can detect precise location of points of interest at subpixel level, thus allowing for a lower resolution of the camera [26, 27], further increase the feasibility of incorporating vision sensor technology in low-power and small-form-factor wearable smart sensors. Likewise, the SOA in IMU technology has reached such a point that open-source data fusion algorithms can provide accurate and precise orientation measurements [1]. These advances in the vision and IMUs create a need for novel multimodal sensor fusion algorithms and system architectures to utilize these emerging possibilities. Whereas the IS-1500 is a very accurate inside-out motion tracker, it is complex and has high overall requirements. On the other hand, the two outside-in trackers, proposed by Maereg et al. and Li et al., offer less expensive and less complex alternatives. However, they are both outside-in trackers that require externally mounted camera with the two points of reference having to be attached to the object of interest which may not be as practical as the inside-out approach.

The proposed system that this work describes aims at combining the strengths of the IS-1500 and the trackers proposed by Maereg et al. and Li et al. and overcoming their limitations. Therefore, the proposed system is an opto-inertial, inside-out, tracker, like the IS-1500, but it requires only two external reference points, like the systems proposed by Maereg et al. and Li et al., and it can be implemented as an embedded wearable, motion tracking system. Although it can be used for motion tracking in various application spaces, this work focuses on ST and the correctness of executing the barbell squats as a specific use case. The subsequent sections describe it in detail.

III. Methodology

This section describes the proposed system architecture and the multimodal sensor fusion algorithm that leverages its properties. First, the overall system is described in detail. It includes the hardware specifications. Subsequently, the proposed algorithm is validated in several simulated and experimental scenarios.

A. System Architecture

Human motion tracking using wearable smart sensors requires a thoughtful consideration of many factors, especially in the context of applications that require low-power and small form factor. The proposed WP incorporates a monocular vision sensor, which can have negative implications on the performance. Despite its advantages, vision sensors require a considerable amount of computational power to process multiple Frames Per Second (FPS), each with many pixels; often counted in millions, i.e. Mega Pixels (MP). The WP needs to be able to process the image frames at a relatively high frame rate; in tens of FPS. Furthermore, the type of information that needs to be extracted from the image frames has a

XXXX-XXXX © XXXX IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

significant impact on the complexity of the image processing algorithms used in this task. For example, a high noise floor in the images, accompanied by the complexity of the points of interest to be found, can dramatically increase the computational requirements of the system. Hence, a human motion tracking system in this context needs to consider all factors; including the software/firmware, hardware as well as the ambient environment beyond the WP.

The proposed system can be broken down into two main elements, the WP, and the Ambient Environment, as shown in Fig. 1. These two elements are connected together via a Radio Frequency (RF) telecommunications link. The RF link enables an interaction between these two elements to help ensure that the system operates in its optimum conditions. The optimum conditions, in this context, are defined as such that the camera can reliably capture the two points of reference and the point tracking image processing algorithm can accurately determine the centres of these points. To this end, the light intensity of the IR LEDs can be continuously adjusted, as the WP moves in the 3-D space. Therefore, the WP can communicate with the Ambient Environment and send commands to adjust the intensities of the IR LEDs so as to make sure that the pixel intensities of the corresponding points in the captured images remained within a specific intensity range. The WP incorporates a monocular vision system and an IMU to perform the inside-out tracking. It also has a Micro Controller Unit (MCU) for data processing, power management block and an RF module.

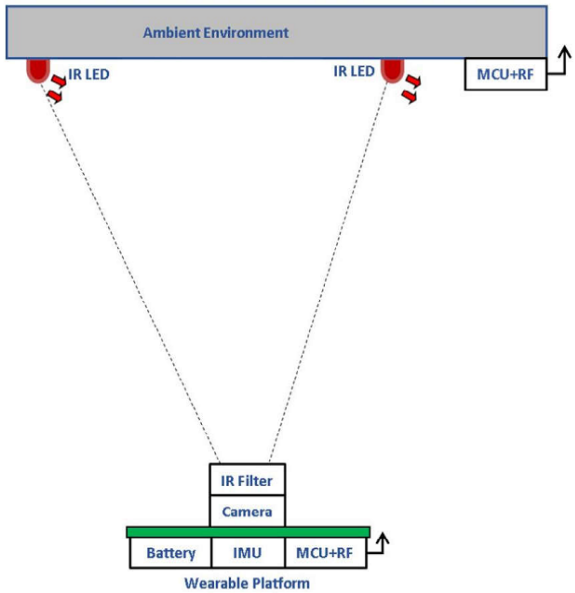

Fig. 1. Generalised system architecture

The Ambient Environment consists of an RF module with an MCU and the two points of reference. This system was designed with active markers as the points of interest to be tracked. The Infrared (IR) LEDs are tracked by the camera in the WP, which has a matching IR filter attached to it. The Ambient Environment includes two IR LEDs and a control unit to maintain the optimum conditions for WP.

The optimum conditions of the system are such that the intensities of the two IR LEDs are set so as to ensure that their pixel intensity profiles, measured by the camera, are in the optimum range in all 3-D poses of the WP; i.e. neither too high (no saturated pixels) nor too low (point peaks are not buried in the noise floor). It is important because changes in position and/or orientation of the WP cause changes in intensities and dimensions of the IR LEDs as captured on the camera's pixel array, which in turn can have a negative effect on the performance of the point detection and tracking tasks. Therefore, these changes are to be offset by controlling the intensity of the IR LEDs. In practical terms, the intensity is maintained in the range of values between 33 % and 66 % of the maximum intensity. To this end, the WP communicates with the Ambient Environment and sends commands to adjust intensities of the IR LEDs. The MCU in the Ambient Environment, in turn, adjusts the intensities accordingly, by driving the IR LEDs with Pulse Width Modulation (PWM). The WP can send a command to increase or decrease intensities of the two IR LEDs after each image frame was processed. The intensity range of the LEDs is mapped to the 8-bit pixel intensity range of the camera in the WP. For example, the pixel intensity of 85 on the pixel array, on the camera in WP, corresponds to approximately 33 % of the LED's intensity range. Likewise, 170 corresponds to approximately 66 % of the LED's intensity.

The work envelope, i.e. the space in which the WP can operate, of the system was designed with simplicity and scalability in mind. Since the proposed 3-D pose detection algorithm relies on two points of reference in the ambient environment, details of which will be described in the following section, the two IR LEDs must be within the FoV of the WP's camera. Also, given the fact that most of ST exercises are stationary, the work envelope does not need to be large. Though, it needs to be scalable. As a result, the work envelope for the system was designed with an arbitrarily set distance between the IR LEDs, called the baseline $B = 500\ mm$. This value of B allows for meeting two objectives. Firstly, the WP can perform translation within the work envelope with a relatively wide range of rotations, while retaining both reference points in the FoV of the camera. Secondly, the

XXXX-XXXX © XXXX IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

calculations in the proposed algorithm yield more accurate results if the distance between the two IR LEDs, as captured by the camera, is relatively large. The reason for it is that the proposed algorithm relies on the geometries formed in the system, which is described in detail in the following section.

The size of the work envelope can be scaled, up or down, by adding additional IR LEDs separated by the baseline distance B. The IR LEDs can be switched ON and OFF using the RF link; as the WP changes its position. The size of the work envelope can also be changed by varying the value of the baseline B. However, the scope of this work is to describe the fundamental principles of this system and prove the accuracy of the algorithm developed to work on the resource constrained WP processor. Thus, the use of two IR LEDs with a fixed B is described in this work. The work envelope is shown in Fig. 2. It is effectively a 3-D space whose boundaries are defined by the continuous, thick, line segments. Its dimensions have a twofold impact on the system. Firstly, the intensity of the LEDs can be controlled dynamically to maintain the optimum level for the camera in the WP. Secondly, both reference points remain within the FoV of the camera; with the exception for certain orientations in the boundary regions. These parameters match the requirements of our target application space; especially that of certain ST exercises, such as the barbell squat.

It needs to be noted that the naming conventions from robotics engineering were adopted in this work. The right-handed coordinate system was used. The origin of the global, or World, coordinate frame L_W is coincident with the location of the reference point P_0^W (read as point zero in World reference frame), as shown in Fig. 2. The two IR LEDs were located $1000\ mm$ above the ground (to match the conditions in our laboratory), thus placing the origin of frame L_W at that height.

B. 3-D Pose Detection Algorithm

Fig. 3 shows the general block diagram of the proposed sensor fusion algorithm. The Data Fusion block is where the 3-D pose is computed. It takes in three inputs: the coordinates of the two reference points extracted from the image frame, expressed in Image frame, $p^I = [p_0^I\ p_1^I]^T$, the orientation of the WP from the IMU, expressed in the World frame of reference, $\theta^W = \left[\theta_x^W\ \theta_y^W\ \theta_z^W\right]^T$, and the camera intrinsic calibration parameters.

The orientation of the WP in the World frame of reference, i.e. the vector θ^W, can be obtained by transforming the IMU's output orientation to World frame of reference.

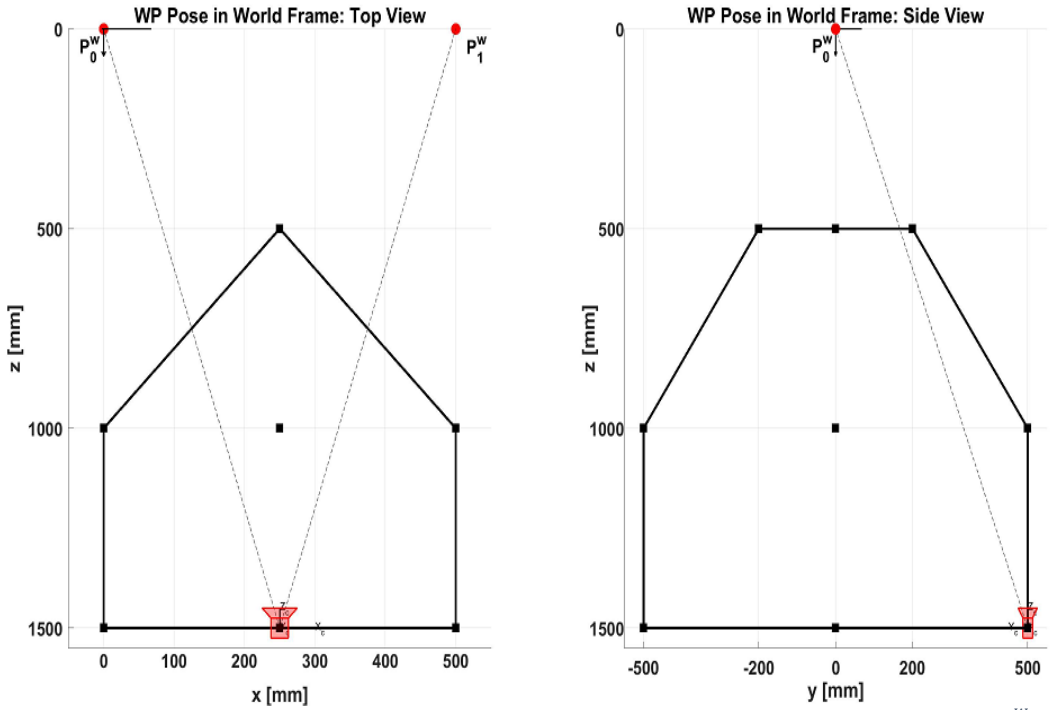

Fig.2. Wearable platform (represented by the camera symbol) inside the work envelope (thick continuous line) with reference points P_0^W and P_1^W (IR LEDs) in camera's FOV

XXXX-XXXX © XXXX IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

If the IMU is calibrated correctly, Madgwick's algorithm returns orientation in Earth's frame of reference L_E; defined by Earth's magnetic and gravitational fields [1]. Therefore, the homogenous transformation matrix from Earth, L_E, to World, L_W, frame of reference, T_W^E, can be defined one containing a the rotation matrix with the translation elements set to zero. In practice, the y-axes in frames L_W and L_E are parallel to each other, i.e. $\hat{y}^W \parallel \hat{y}^E$, and can be assumed to be pointing in the same direction, i.e. their dot product is $\hat{y}^W.\hat{y}^E = 1$. Therefore, the transform T_W^E is reduced to describing a fixed rotation about the $\hat{y}^W$- axis.

This transformation is then used for transforming the orientation of WP from L_C to L_W, as follows. The vector of rotation angles of the WP measured by the IMU, $\theta^E = \left[\theta_x^E\ \theta_y^E\ \theta_z^E\right]^T$, can be also represented as a homogenous transformation matrix from Camera, L_C, to Earth, L_E, frame of reference T_E^C; with the X-Y-Z order of rotations in the rotation elements and the translation elements set to zero [28]. Therefore, the transformation from L_C to L_W, i.e. T_W^C, is defined as shown in (1). Subsequently, the rotation angles of the orientation vector θ^W can be extracted from it.

$$T_W^C = T_W^E T_E^C \tag{1}$$

The intrinsic camera parameters for the specific vision sensor can be calculated via a camera calibration process [29]. The intrinsic parameters, along with the knowledge of the specific image sensor from its datasheet, such as the focal length f, pixel dimension and size and location of the optical centre, are used to transform p^I to the Camera reference frame L_C expressed in metric units; resulting in p^C. The output is the 3-D pose of the WP defined as the position and orientation in the World frame of reference as follows $P_{WP}^W = [P\ \theta]^T = \left[P_x^W\ P_y^W\ P_z^W\ \theta_x^W\ \theta_y^W\ \theta_z^W\right]^T$. The subscripts in the variables define the axis. For example, the angle θ_x^W is the rotation angle about the $\hat{x}$-axis in the World frame of reference. Note, the hat symbol implies the axis component of a unit vector, e.g. $\hat{x}^W$ means the $\hat{x}$-axis in World reference frame. The proposed data fusion algorithm computes the 3-D pose in three discrete steps, as shown in Fig. 4.

- Step 1 corrects the input points p^I using the rotation angle of the WP about the $\hat{z}$-axis in World frame θ_z^W. The subsequent two steps break down the problem into two smaller tasks.
- In Step 2, the position P_x^W is computed on the $\hat{x}^W\ \hat{z}^W$-plane with θ_y^W.
- In Step 3, the position elements P_z^W and P_y^W are computed on the $\hat{y}^W\ \hat{z}^W$ −plane and with θ_x^W, to finally yield the result, i.e. the 3-D pose of the WP in the World frame of reference P_{WP}^W.

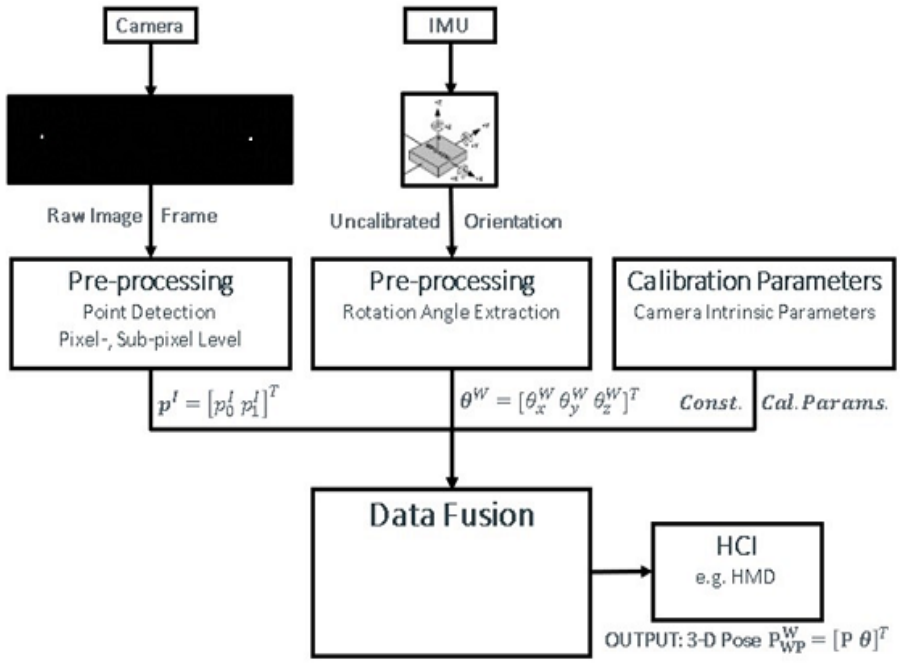

Fig. 3. General block diagram of the proposed data fusion system (raw input frame contains two points of reference)

XXXX-XXXX © XXXX IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

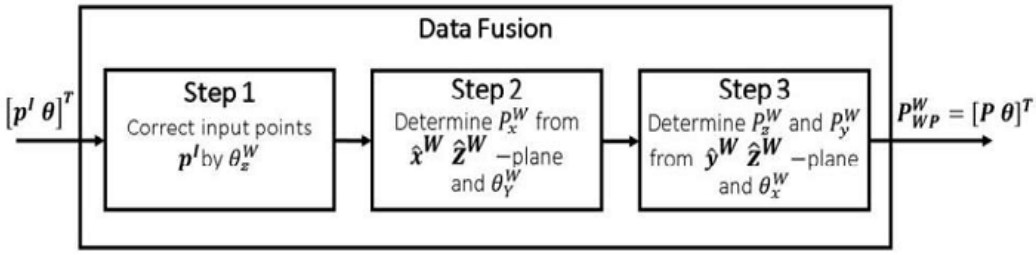

Fig. 4. Block diagram of the proposed data fusion algorithm

The three steps of the proposed algorithm are described in detail in the subsections below.

1) *Step 1 – Input Points Correction*

The geometric model that is used in calculating the pose of the WP achieves the best results when the $\hat{x}$-axis or the $\hat{y}$-axis of the L_C, i.e. that of the WP, and L_W reference frames are parallel, or close to it. It is so because the calculations in Steps 2 and 3 of the proposed algorithm are carried out on the planes $\hat{x}^W\hat{z}^W$ and $\hat{y}^W\ \hat{z}^W$, respectively.

In other words, the calculations are more accurate if the rotation matrix from L_C to L_W reference frame R_W^C is as close as possible to that defined in (2). This condition means that all corresponding axes are parallel; with $\hat{y}$-axes and $\hat{z}$-axes of these two reference frames pointing in opposite directions. It simplifies the geometry formed by the IR LEDs and the camera. Effectively, the line segment between points p_0^I and p_1^I extracted from the image frames needs to be parallel with the $\hat{x}$-axis of the frame L_C. However, it is not a realistic scenario. It effectively makes the WP's orientation constant, such that it directly faces the IR LEDs, with only the translation being allowed to vary. It is obviously an unacceptable condition in the context of the considered application space. Therefore, our algorithm uses a corrective step to meet this condition, or at least approximately match it.

$$R_W^C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \tag{2}$$

The corrective step is applied to point p^C. Whereas it would be a straightforward process in 3-D, it is more complicated in the case of the two points p^C. In the case of 3-D points the data from the calibrated IMU could be used to rotate the points. However, the translation vector of the WP P is unknown at this step. In fact, the objective of this work is to determine P.

The proposed solution to this problem takes advantage of the fact that many ST exercises are largely stationary with a predefined body posture and range of motion. For example, a barbell squat would involve relatively little rotation and some translation if the WP was attached to the back of the exercising individual. From a technical point of view, it means that the WP would face the reference points in the ambient environment. It needs to be noted that the initial rotation matrix R_W^C is the same as that defined in (2). Also, the rotation angles would be relatively small. Hence, our corrective step involves a two-dimensional rotation of the image points $p^C$by rotation angle θ_z^C, as defined in (3). This angle is not negative, because we are correcting the orientation of the WP. The operation or rotating points p^C by θ_z^C, which is effectively θ_z^W, is an approximate equivalent of rotating the WP in the opposite direction.

$$rRot(\hat{z}^C, \theta_z^C) = \begin{bmatrix} cos\,\theta_z^C & -sin\,\theta_z^C \\ sin\,\theta_z^C & cos\,\theta_z^C \end{bmatrix} \tag{3}$$

Subsequently, the two transformed points p^C are passed to Step 2 in the algorithm.

Step 2 – Calculation of P_x^W

In this step, the position P_x^W of the WP is computed using the $\hat{x}^W\ \hat{z}^W$-plane, as shown in Fig. 5. The $\hat{y}$-axis is ignored in this step because the algorithm performs the calculation only on the $\hat{x}^W\ \hat{z}^W$-plane. The elements of the general system architecture, shown in Fig. 1, directly correspond to the geometric model shown in Fig. 5. The IR LEDs correspond to the points P_0^W and P_1^W while the camera is expressed as the large rectangle. The IR Filter and Lens are assumed to be ideal elements that don not affect the system to simplify the model.

This model enables the calculation of the 3-D pose due to its specifically designed architecture. Firstly, the baseline B is known. Secondly, the camera's intrinsic parameters can be determined by camera calibration. The camera calibration routine can determine the key parameter of the camera that is critical in the calculations, i.e. the focal length f. Furthermore, the knowledge of these parameters, complemented with the rotation angles from the IMU, enabled us to use geometry and trigonometry to compute the pose.

XXXX-XXXX © XXXX IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

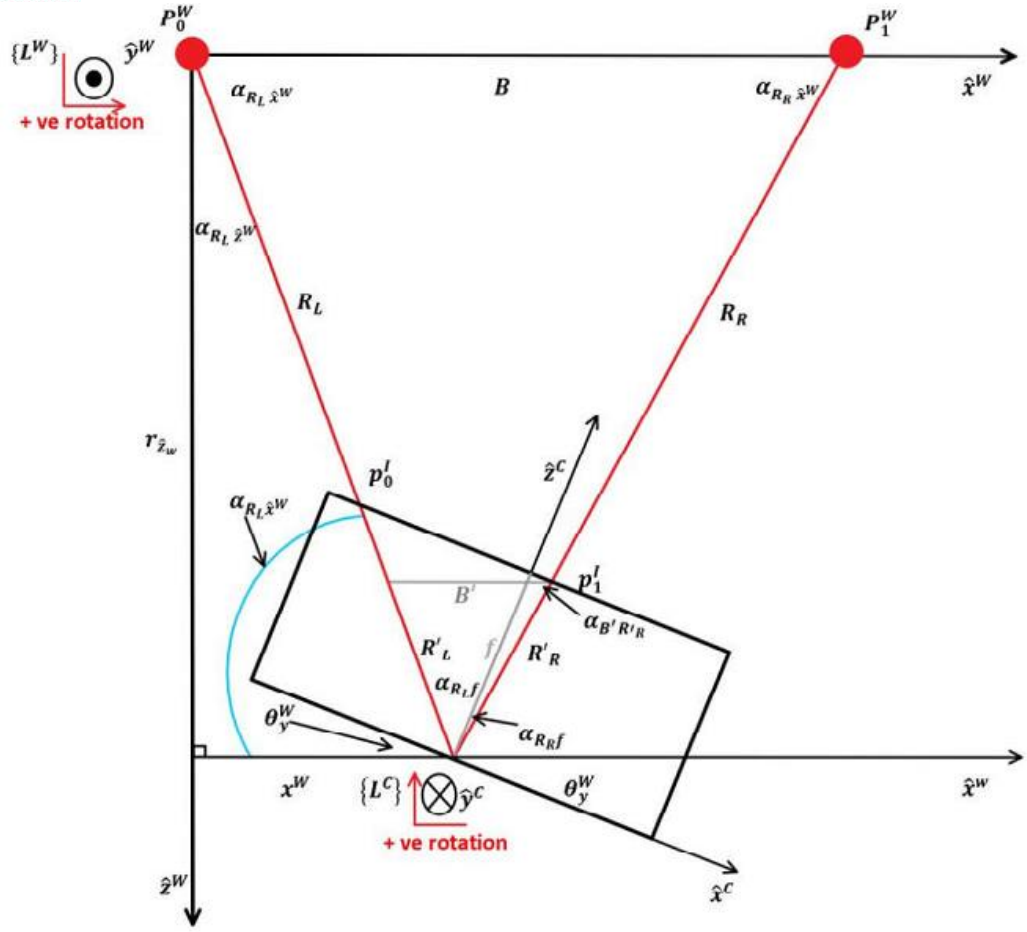

Fig. 5. Geometric model of the system, $\hat{x}^W - \hat{z}^W$-plane

The knowledge about the orientation of the WP makes it possible to use geometry to solve our problem. The properties of similar triangles and trigonometry are particularly useful.

The camera can be modeled with a simplified projection model, i.e. one in which the image plane is in front of the principal point, which is coincident with the origin of the Camera frame L_C; as opposed to being behind it. The image points p_0^I and p_1^I are the projections of their corresponding World points P_0^W and P_1^W on the camera's image plane. The two rays of light, R_L and R_R, that originate from the two World points and pass through their corresponding Image points intersect at the L_C. The rotation angles from the IMU help us form two similar triangles. The first triangle has the following vertices P_0^W, P_1^W, and L_C. The second triangle has the following vertices p_0^I, p_1^I, and L_C. The image points p_0^I, p_1^I are transformed to the Camera frame to enable real-world-unit calculations, i.e. p_0^C, p_1^C. The proportions are achieved by making B and B' parallel.

The first task in this step is to compute the angles: between the left light ray R_L and the line segment of length equal to the focal length f, angle between R_R and f, angle between the R_L and f, angle between R_L and the axis $\hat{x}^W$, angle between R_L and the axis $\hat{z}^W$, angle between the rays R_L and R_R, and the angle between B' and R'_R; defined in: (4), (5), (6), (7), (8), and (9), respectively.

$$\alpha_{R_L f} = tan^{-1}\left(\frac{p_0^C}{f}\right) \tag{4}$$

$$\alpha_{R_R f} = tan^{-1}\left(\frac{p_1^C}{f}\right) \tag{5}$$

$$\alpha_{R_L \hat{x}^W} = \frac{\pi}{2} + \alpha_{R_L f} - \theta_y^W \tag{6}$$

$$\alpha_{R_L \hat{z}^W} = \frac{\pi}{2} - \alpha_{R_L x^W} \tag{7}$$

$$\alpha_{R_L R_R} = \left(\alpha_{R_R f} - \alpha_{R_L f}\right) \tag{8}$$

XXXX-XXXX © XXXX IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

$$\alpha_{B'R'_R} = \pi - \alpha_{R_L\hat{x}^W} - \alpha_{R_LR_R} \tag{9}$$

The length of the line segment R'_L is calculated with (10), which then allows us to determine the value of B' using (12), using the sine rule and transposing (11).

$$R'_L = \sqrt{f^2 + {p_0^c}^2} \tag{10}$$

$$\frac{B'}{sin\,(\alpha_{R_LR_R})} = \frac{R'_L}{sin\,(\alpha_{B'R'_R})} \tag{11}$$

$$\therefore B' = \frac{R'_L sin\,(\alpha_{R_LR_R})}{sin\,(\alpha_{B'R'_R})} \tag{12}$$

The properties of Similar Triangles can be used to find the length of R_L with (13) followed by (14).

$$\frac{R_L}{R'_L} = \frac{B}{B'} \tag{13}$$

$$\therefore R_L = R'_L\left(\frac{B}{B'}\right) \tag{14}$$

In the final stage, the trigonometry is used to find the values of the remaining two variables. The sine function is used to find x^W with (15) and (16), which is in effect equal to one of the elements of the 3-D Pose P_x^W.

$$\frac{x^W}{R_L} = sin\left(\alpha_{R_L\hat{z}^W}\right) \tag{15}$$

$$\therefore x^W = R_L sin\left(\alpha_{R_L\hat{z}^W}\right) \tag{16}$$

Finally, the value of $r_{\hat{z}_W}$ is computed using the cosine function with (17) and (18). The radius $r_{\hat{z}_W}$ is required in the computations in Step 3.

$$\frac{r_{\hat{z}^W}}{R_L} = cos\left(\alpha_{R_L\hat{z}^W}\right) \tag{17}$$

$$\therefore r_{\hat{z}^W} = R_L cos\left(\alpha_{R_L\hat{z}^W}\right) \tag{18}$$

2) *Step 3 – Calculation of P_z^W and P_y^W*

The remaining two unknown variables are computed in this step, i.e. the y^W and z^W. The y^W and z^W correspond to the P_y^W and P_z^W elements of the P_{WP}^W vector, respectively. The computations are carried out on the $\hat{y}^W\,\hat{z}^W$ −plane. The geometric model of the system is shown in Fig. 6. The corrective rotation that was applied in Step 1 lets us assume that the axes of the frames L_C and L_W are approximately aligned with the rotation transformation R_W^C close to that defined in (2).

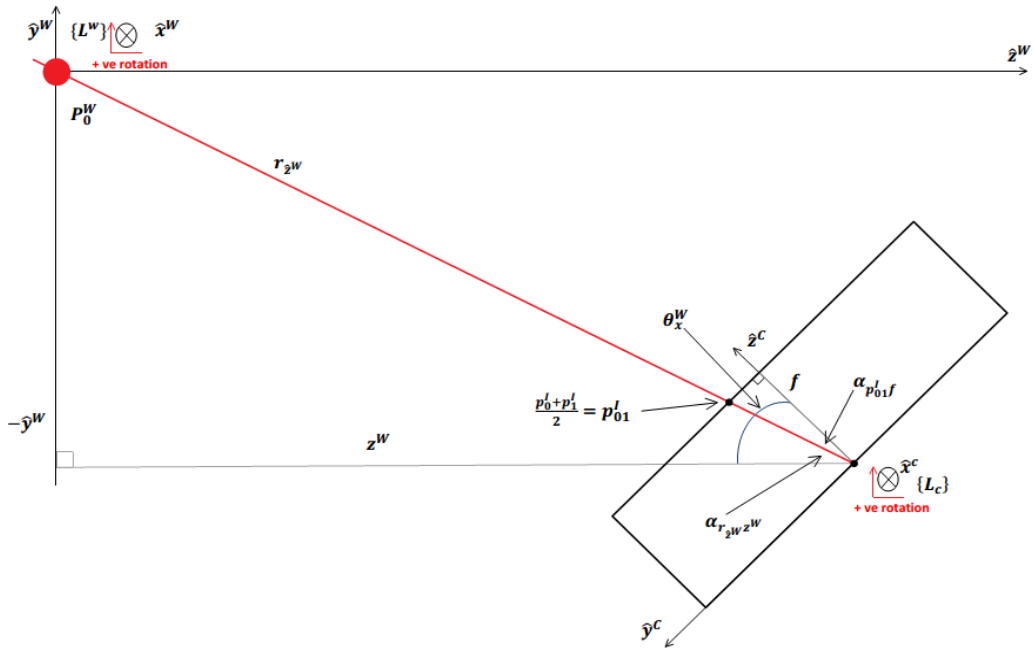

Fig. 6. Geometric model of the system, $\hat{y}^W - \hat{z}^W$-plane

XXXX-XXXX © XXXX IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

As in the previous step, the system setup allows us to use trigonometry to determine the missing pieces of information. It is effectively the side-view of the system. The calculations use three inputs. Given the corrections described in Step 1, the line segment formed by the image point vector p^I is effectively parallel to $\hat{x}^W$, correct to approximately within $1\,deg$. The mid-point between these two points p_{01}^I is used; specifically, the vertical coordinate on the image plane. As in the previous step, the p_{01}^I is transformed to p_{01}^C for calculations in real-world-units. Also, the $\hat{x}^W$-axis is ignored in this step.

The angle $\alpha_{p_{01f}^C}$ is found using the right-angled triangle with vertices at: intersection of $\hat{z}^C$ with image plane, the mid-point p_{01}^C, and the origin L_C. Thus, the inverse tangent of of the ration of the $p_{01}^C$to the focal length f is equal to this angle, as defined in (19). The angle between the $\hat{z}^W$-axis light-ray R_L, whose length is $r_{\hat{z}}$w, is found by correcting $\alpha_{p_{01f}^C}$ by θ_x^W, as shown in (20). Finally the remaining unknowns z^W and y^W are found using cosine and the negative sinus functions of $\alpha_{r_{\hat{z}W}\hat{z}W}$, scaled by $r_{\hat{z}w}$, defined in (21) and (22), respectively.

$$\alpha_{p_{01f}^C} = tan^{-1}\left(\frac{p_{01}^C}{f}\right) \tag{19}$$

$$\alpha_{r_{\hat{z}W}\hat{z}W} = \theta_x^W - \alpha_{p_{01f}^C} \tag{20}$$

$$z^W = r_{\hat{z}w} cos\left(\alpha_{r_{\hat{z}_w}\hat{z}W}\right) \tag{21}$$

$$y^W = -\,r_{\hat{z}w} sin\left(\alpha_{r_{\hat{z}_w}\hat{z}W}\right) \tag{22}$$

At this point the 3-D Pose is computed. The elements of the pose vector are as follows: $P_{WP}^W = [P\ \theta]^T = \left[P_x^W\ P_y^W\ P_z^W\ \theta_x^W\ \theta_y^W\ \theta_z^W\right]^T = \left[x^W\ y^W\ z^W\ \theta_x^C\ -\theta_y^C\ -\theta_z^C\right]^T$. The orientation angles θ, measured by the IMU, determine the orientation of the WP. The orientation of the WP in World and Camera frame are the same, with the exception for the signs of some of its elements; due to the fact that WP faces the IR LEDs, and the rotation matrix R_W^C is assumed to be relatively close to that defined in (2).

C. System Modelling

The proposed system, along with the sensor fusion algorithm described in the previous section, was modeled, and evaluated in simulated conditions. The objective of this task was twofold. Firstly, the system's performance was to be simulated in a number of scenarios. Secondly, the impact of various noise levels originating from uncertainties in point detection and orientation estimation processes was to be determined. The proposed system was modeled and evaluated in MATLAB®.

One of the key elements in modelling the system was the camera. To be able to simulate it in a realistic way, the camera had to be carefully modeled. The locations of the two input points of reference p_0^I and p_1^I, captured by the camera, as visualised in Fig. 3, had to closely correspond to their respective locations in World frame, as shown in Fig. 2. This correspondence was critical in achieving the ability to compare the results calculated by the proposed system to the real-world position and orientation of the WP. The pinhole camera model is commonly used to map 3-D World points to 2-D Image points, given the intrinsic, extrinsic, and distortion parameters of the camera, in applications that such as camera calibration. In this work, we used a MATLAB® implementation of this model developed by Zachary Taylor [30]. It was used for projecting 3-D points onto a 2-D image plane using camera calibration parameters, the 3-D coordinates of the two reference points, P_0^W and P_1^W, and the extrinsic matrix. The camera calibration parameters were obtained from the same camera module that was used in the experimental work (described in Section II D). Likewise, the focal length f, which was required by the proposed algorithm, was obtained from the intrinsic matrix. The extrinsic matrix is a transform that describes pose of the WP in World frame of reference. Thus, the input position and orientation of the WP in World frame of reference was encoded in this transform matrix and passed to the function that projected the two 3-D reference points and output the 2-D image points. The two Image points were subsequently used as one of the two inputs to the proposed data fusion algorithm. The second input was the orientation vector θ^W, which was also used in constructing the extrinsic matrix.

1) *Evaluation Scenarios*

The proposed system was evaluated in several scenarios. In each case, $N > 5000$ appropriate inputs were generated and passed to the data fusion algorithm. The following scenarios were used in this process:

- Scenario 1 - Linear motion – along $\hat{x}\hat{y}\hat{z} - axes$

In this scenario, the WP moved on a straight line across the Work Envelope along all three axes, i.e. $\hat{x} - \hat{y} - \hat{z}$ in World frame of reference. The translation along the axes was as follows: $x^W \in < 150, 350 > mm$, $y^W \in < -250, 250 > mm$, $z^W \in < 1000, 1500 > mm$. The orientation vector was set to $\theta^W = [0, 0, 0]^T$ deg, and it did not vary.

- Scenario 2 – Uniform Random

In this scenario, the proposed system was evaluated under the most challenging conditions. The position and

orientation of the WP were varied at random; using random number generator with uniform probability distribution. All elements of the pose vector of the WP were varied simultaneously. The range of possible positions and orientations were set such that the system was evaluated under all possible poses, including the extreme ones near the edges of the Work Envelope. The position and orientation ranges were set as follows: $x^W \in < 0, 250 > mm$, $y^W \in < 0, 250 > mm$, $Z^W \in < 1000, 1500 >$ mm, $\theta_x^W \in < 0, -10 > deg$, $\theta_y^W \in < 0, -10 > deg$, $\theta_z^W \in < 0, 10 > deg$. Although the range of positions covers only 25 % of the Work envelope for $z^w \in < 500, 1500 > mm$, it is safe to expect similar performance across the remaining volume in this range of z^W as it is a symmetrical system. It needs to be noted, that a check was performed for each pose in this scenario to ensure that both points of reference were present in camera's FoV, which was the prerequisite for the proposed data fusion algorithm to work. This condition was possible for such poses that $z^W \in < 500,\ 1000 >$ mm and the magnitude of the other elements of the pose vector of the WP were close to their maximum values in their respective ranges.

- Scenario 3 – Linear motion – along $\hat{y} - axis$

In this scenario, the WP moved on a straight line across the Work Envelope along all the $\hat{y} -$ axis in World frame of reference. The translation was as follows: $x^W = 250\,mm$, $y^W \in < -500, 300 > mm$, $z^W = 1400\,mm$. The orientation vector was set to $\theta^W = [0, 0, 0]^T$ deg, and it did not vary.

This scenario was of most interest to this work. It was designed to simulate the pattern of motion involved in the barbell squat with the correct technique. It was assumed the WP was attached to the back of the person executing the exercise, e.g. under the bar, between upper and lower back. In this case, there would not be much rotation expected about any axis [10]. The motion would be largely vertical with full range of motion, i.e. parallel squat, with little lateral hip shift or trunk lean [6, 9, 11].

2) *Point and IMU Noise*

The proposed data fusion algorithm is susceptible to noise that is expected to be present in the input position and orientation vectors, p^I and θ^W, respectively. The individual sources of error as well as their magnitude have a negative impact on the system. This subsection describes the process of quantifying it.

The point noise, i.e. error in the coordinates of the image points in the p^I vector, may originate from several sources. One of the most common causes are the imperfections in the camera, which were not sufficiently rectified by the camera calibration process. For example, the lens distortions may significantly alter the coordinates of points on the image plane; especially at larger distances between those points and the optical centre on the image plane. The accuracy of point detection algorithms may also be affected if the angle between the optical axis of the camera and the line segment between its optical centre and the point of interest increases. Under these conditions, the shape of IR LED may resemble an ellipsoid on the pixel array, instead of a circle. The level of point noise may be measured in pixels. Its magnitude generally depends on the pixel resolution of the camera and where on the image plane the points were captured. The angles the camera was at during image capture, relative to the given point, plays are role, too. Several empirical tests were carried out to determine the maximum level of point noise using the same camera module as that used in the calibration and experimental work (described in Section II D). The tests showed that the point noise was generally bounded to 10 pixels. As a result, point noise was modeled as a Gaussian noise distribution $\mathcal{N}_P(\mu_P, \sigma_P)$ with mean μ_P set to the noise-free input vector p^I for the given scenario and maximum standard deviation σ_P, thus resulting in p^I containing the added point noise. The maximum standard deviation was set to $\sigma_P = 10\,pixels$.

The IMU noise considered in this work was defined as the error in orientation angles of the WP, i.e. the vector θ^W. This noise may have numerous sources, ranging from poor IMU calibration to suboptimal configuration or the sensor fusion algorithm. Nevertheless, the error in orientation estimation, computed by sensor fusion algorithms, is generally bounded to $1\,deg$, [1]. Similarly to the point noise, the IMU noise was modeled with a Gaussian noise distribution $\mathcal{N}_{IMU}(\mu_{IMU}, \sigma_{IMU})$ with the mean μ_{IMU} being set to the noise-free input vector θ^W for the given scenario and the standard deviation σ_{IMU}, thus resulting in θ^W containing the added IMU noise. The maximum standard deviation was set to $\sigma_{IMU} = 1\,deg$.

The performance of the proposed system was evaluated by subjecting it to both noise types in each of the simulated scenarios. The level of noise was increased incrementally. In each scenario, the system was subjected to five different levels of noise, which was defined as a vector $\mathcal{N}_i = [\sigma_{Pi};\ \sigma_{IMUi}]$, where $\sigma_{Pi} = [0, 2.5, 5, 7.5, 10;]pixels$ and $\sigma_{IMUi} = [\,0, 0.25, 0.5, 0.75, 1;\,]\,deg$. At each level of noise $\mathcal{N}_i$, three different combinations of this noise were applied to the system: $\mathcal{N}_{IMU}$ only, $\mathcal{N}_P$ only, both $\mathcal{N}_{IMU}$ and $\mathcal{N}_P$. Thus, the individual and combined impact of noise could be examined. Note, the case with no added noise was examined at $i = 0$, i.e. $\mathcal{N}_0 = [0;\ 0;]$.

XXXX-XXXX © XXXX IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

3) *Error Analysis in 3-D Positioning*

The main performance metric in measurement accuracy was the RMSE in 3-D, as defined in (23). The algorithm's output is defined as d_i and the corresponding reference values as $\hat{d}_i$ over all N measurements in this equation.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\hat{d}_i - d_i\right)^2} \qquad (23)$$

The RMSE was computed for each: scenario, noise level and noise source combination. The RMSE was determined for each position element of vector P_{WP}^{W}, as well as the combined error over all three axes. The results for scenarios 1, 2 and 3 are shown in figures Fig. 7, Fig. 8, and Fig. 9, respectively. It can be seen that RMSE increased in all three scenarios with the increase in noise level $\mathcal{N}_i$. The IMU noise $\mathcal{N}_{IMU}$, in most cases, has a greater impact on the RMSE than the point noise $\mathcal{N}_P$. Due to the random distribution of both noise sources, $\mathcal{N}_{IMU}$ and $\mathcal{N}_P$, the RMSE was lower than the sum of the individual RMSE values when both noise sources were applied to the system, i.e. both $\mathcal{N}_{IMU}$ and $\mathcal{N}_P$; as compared to the conditions with noise sources applied separately, i.e. either $\mathcal{N}_{IMU}$ or $\mathcal{N}_P$.

The system achieved the lowest RMSE in scenario 1. Although, the position of WP varied across all three axes, the range of motion was relatively small, thus avoiding the unfavourable conditions. On the other hand, scenario 3 was the most challenging one. It was designed to determine the performance in the most adverse conditions under which it the proposed system could still perform without failing. The system would fail if any one of the two reference points was outside the camera's FoV, or the intensity of the IR LEDs was too low for the camera to capture. As a result, the RMSE was the highest in this case. Nevertheless, the RMSE was not significantly higher in this scenario, as compared to scenario 1.

The total RMSE can be broken down into individual components, i.e. the errors in the three position elements P_x^W, P_y^W, P_z^W of the pose vector P_{WP}^{W}. The analysis can show that the RMSE was not equally distributed across these three position elements, and it depended on the noise level $\mathcal{N}_i$. The RMSE on all axes for different levels of noise $\mathcal{N}_i$ is shown in

TABLE *I*. The RMSE in P_z^W was the largest component of the total RMSE. Its value was the closest to the overall RMSE whereas RMSE in P_x^W and P_y^W was significantly lower. The RMSE in P_y^W was much higher than that in P_x^W at low values of $\mathcal{N}_i$.

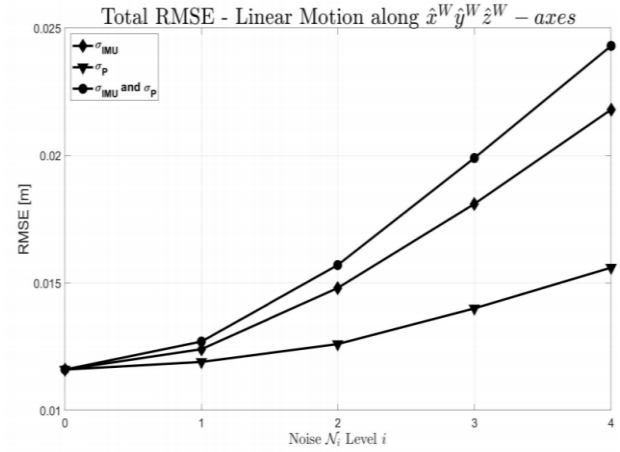

Fig. 7. RMSE in Scenario 1 - Linear Motion along $\hat{x}^W\hat{y}^W\hat{z}^W$- axes for different levels of noise $\mathcal{N}_i$

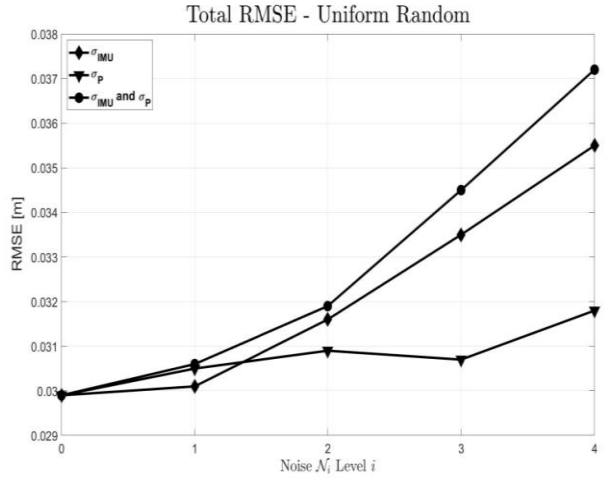

Fig. 8. RMSE in Scenario 2 – Uniform Random for different levels of noise $\mathcal{N}_i$

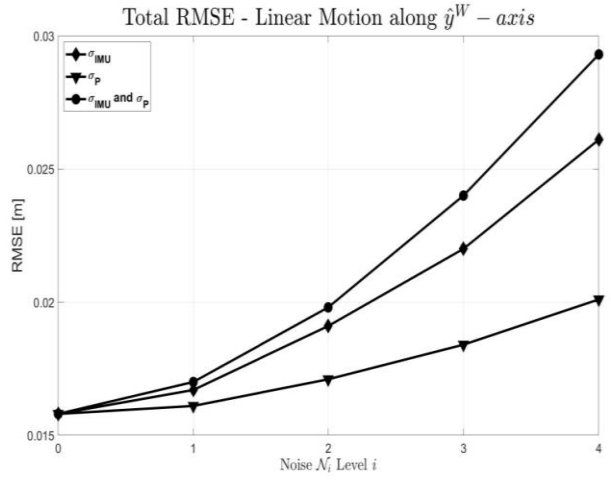

Fig. 9. RMSE in Scenario 3 - Linear Motion along $\hat{y}^W$- axis for different levels of noise $\mathcal{N}_i$

XXXX-XXXX © XXXX IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

The difference in RMSE between P_x^W and P_y^W decreased with increasing values of noise $\mathcal{N}_i$. The RMSE in P_y^W was approximately 50 % lower than that in P_z^W at low noise level $\mathcal{N}_i$ and approached as the noise increased.

The visual representation of position computation in scenario 3 is shown in Fig. 10. This simulation was executed with noise level $\mathcal{N}_1 = [\sigma_{P1};\ \sigma_{IMU1};] = [2.5;\ 0.25;]$ [pixel; deg] to show the impact of added noise. This figure shows visually why the RMSE was lower for P_x^W and P_y^W as compared to P_z^W, as shown in TABLE *I*. Whereas P_z^W deviated away from its reference position as the WP approached the minimum and maximum values of y^W, P_x^W and P_y^W tended to remain close to their corresponding reference values. Thus, the RMSE in P_x^W and P_y^W was relatively low and uniform as compared to RMSE in P_z^W, which was higher and increased near the minimum and maximum values of y^W.

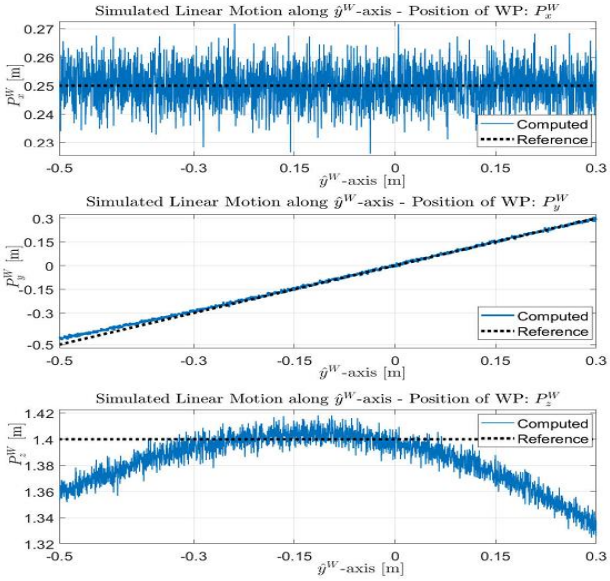

Fig. 10. Simulated Position of the WP in Linear Motion along $\hat{y}^W$-axis with Added Noise $\mathcal{N}_1 = [\sigma_{P1};\ \sigma_{IMU1};] = [2.5; 0.25;]$ [pixel; deg;]

TABLE I
RMSE ON INDIVIDUAL ELEMENTS OF POSE VECTOR P_{WP}^W AND TOTAL RMSE FOR DIFFERENT VALUES OF NOISE $\mathcal{N}_i$ IN SCENARIO 3

$\mathcal{N}_i - i$	RMSE P_x^W [mm]	RMSE P_y^W [mm]	RMSE P_z^W [mm]	Total [mm]
0	0.1	12.6	24.3	15.8
1	6.5	14	25	17
2	12.8	17.7	26.5	19.8
3	19.5	23.1	28.8	24
4	26.3	28.7	32.4	29.3

D. Experimental Work

The proposed system was validated experimentally. The validation process was carried out in two cases, i.e. static and mobile. In the static case, the system was validated in a similar way to that in the simulated scenario 2, i.e. the uniform random. The mobile case closely resembled scenario 3, i.e. the linear motion along the $\hat{y}^W$ axis.

1) *Static Case - Experimental Setup*

The experimental setup corresponded to the general system diagram, shown in Fig. 1, and the work envelope, shown in Fig. 2. The complete implementation of the experimental setup, as described in Section IIA on the system architecture, is shown in Fig. 11.

The WP was implemented using the Microsoft® Surface Pro 4 tablet computer with MATLAB® development environment installed on it. This computing platform was selected due to its portability while being a fully featured computer. Furthermore, it had the built-in OV8865 camera module, which is a low-power camera module, designed for mobile applications. It also featured an MCU unit with a Bluetooth Low Energy (BLE) for control of the IR LEDs. An IMU, the MPU9250 made by TDK InvenSense, was also added to support future functionalities [31]. Additionally, an IR Filter was attached to the camera [32], whose transmittance properties matched the IR LEDs [33], as shown in Fig. 11 (c). The WP was housed in a dedicated, 3-D printed, holder that was mounted on a high-quality camera tripod. The Manfrotto MN755XB aluminium camera tripod with levelling ball with Manfrotto 410 Junior geared head were used in the experiments [34, 35]. The reference pose was measured using a digital protractor and a laser distance meter [36, 37], as shown in Fig. 11 (a).

2) *Static Case - Experimental Data Acquisition*

The input dataset was acquired with the experimental setup, which was at the pre-prototype stage of development. Prior to the acquisition, at each test position the intensities of the IR LEDs were set such that their perceived intensities I on the input image frame's matrix were within the following interval $I(x, y) \in < 63{,}76 >$ on the 8-bit intensity range, which was the optimum intensity for this experimental setup for our subpixel point detection algorithm [26]. Once this condition was met for the given test position, the raw input images were acquired. This process was repeated for each test position in the work envelope marked with squares in Fig. 2, except for those at $x^W > 250\ mm$.

XXXX-XXXX © XXXX IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

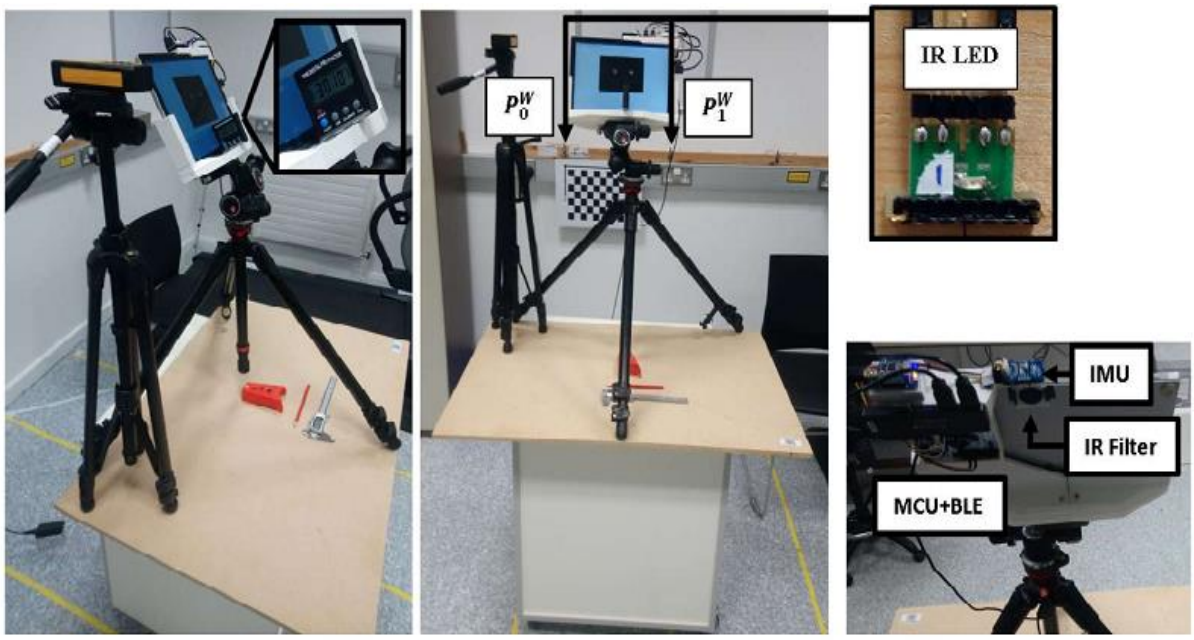

Fig. 11. Experimental Setup – Static Case: (a) Side-View, (b) Front-View, (c) Rear-View

Due to the symmetry along the $\hat{z}^W - axis$, at $x^W = 250\,mm$, it was sufficient to consider only the work envelope with $0 \leq x^W \leq 250\,mm$ and $0 \leq y^W \leq 500\,mm$ ($0 \leq y^W \leq 200\,mm$ at $z^W = 500\,mm$).

A set of ten test positions was selected within this work envelope, with an emphasis on ensuring that all key positions along the external border were included. For each test position, input images were acquired for all orientations, as listed in **Error! Not a valid bookmark self-reference.**. It needs to be noted that some poses at some test positions had to be excluded from the validation. The cases where one or both points of reference were beyond the FoV of the camera invalidated the input frame. At some test positions, the combination of the orientation angle and the position resulted in the camera not being able to capture both reference points. Subsequently, the raw input image frames, along with the corresponding orientation angles, were passed to the sensor fusion algorithm.

3) *Static Case - Results*

The proposed system was evaluated in the experimental laboratory environment. It was experimentally evaluated using the same metric as that used in simulations, i.e. the RMSE. It measured the error in the position estimation along the three axes of the World frame of reference: $\hat{x}^W$, $\hat{y}^W$, and $\hat{z}^W$. The overall RMSE over all three axes combined was also determined; referred to as the Total RMSE, which was the most important metric. The results are shown in

TABLE III.

The RMSE measurement across the individual axes revealed which position elements of the pose were more susceptible to error. It largely confirmed the pattern of noise distribution on the three axes that was present in the simulations. While the position along the $\hat{x}^W$-axis was most accurate, the calculation of the position along the $\hat{z}^W$-axis had the highest RMSE. These results, to some extent, correspond to the simulated scenario 2, i.e. the Uniform Random. Although this scenario did not simulate a static case, the positions and orientations of the WP were similar in both cases. TABLE II

ORIENTATIONS FOR EACH EXPERIMENTAL TEST POSITION

Orientation Index	$\boldsymbol{\theta_x^W}$ $[\boldsymbol{deg}]$	$\boldsymbol{\theta_y^W}$ $[\boldsymbol{deg}]$	$\boldsymbol{\theta_z^W}$ $[\boldsymbol{deg}]$
0	0	0	0
1	-15	0	0
2	-30	0	0
3	0	-15	0
4	0	-30	0
5	0	0	15
6	0	0	30

TABLE III
STATIC CASE - EXPERIMENTAL RESULTS

	$P_x^W[mm]$	$P_y^W[mm]$	$P_z^W[mm]$	Total [mm]
RMSE	17.4	36.7	48.9	36.7

4) *Mobile Case – Experimental Setup*

An experimental setup was designed to validate the performance of the proposed system in a mobile case. The setup was similar to that in the static case shown in Fig. 11. It differed in that the WP was mounted on a motorised mobile track slider system. It enabled the WP to move on a vertical trajectory, along the $\hat{y}^W$-axis in a controlled way, thus closely resembling the simulated scenario 3, which was the main aim of this experiment. Therefore, the position and orientation and range of motion of the WP were the same as those in the simulated scenario 3.

The objective of this experiment was twofold. Firstly, the RMSE was to be determined across the range of $\hat{y}^W$. Secondly, the repeatability of the performance of the proposed system was to be determined. To this end, the WP traversed the distance between y_{min}^W and y_{max}^W twenty times, i.e. it performed ten y_{min}^W-to-y_{max}^W-y_{max}^W-to-y_{min}^W cycles. The mobile experimental setup is shown in Fig. 12.

The track slider was based on the 80 cm version of the Neewer camera slider rail, which was customised for this specific experiment [38]. The slider rail was fitted with a 6-mm-wide T-belt that was connected to the Nema 17 stepper motor via matching 20-tooth pulley wheels [39]. The TB6600 stepper motor was used as the driver for the motor [40]. A Raspberry Pi® computer, Python™ programming environment and Secure Shell connection were used to control motion of the WP from a separate computer. Motion of the WP was controlled with an open-loop motor control system with a trapezoidal velocity profile. The acceleration and deceleration ramps of the velocity profile were set so as to ensure a smooth motion at the inflection points of WP's motion trajectory, i.e. minimum, y_{min}^W, and maximum, y_{max}^W, values of y^W. The maximum velocity was set such that the WP could acquire sufficient amount of input frames to produce statistically significant results. The frame rate of the WP was 30 FPS.

The time the WP required to traverse the distance between y_{min}^W and y_{max}^W, i.e. half a cycle was $\frac{T}{2} = 17$ seconds, where T was the period of one cycle. Given 10 up-down motion cycles, the WP acquired at least 5100 input frames, which was comparable to N samples in simulations.

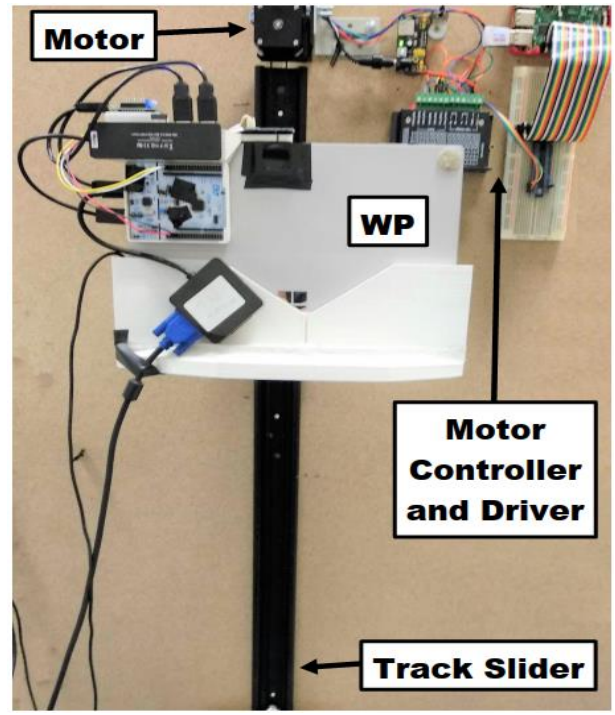

Fig. 12. Experimental Setup – Mobile Case: WP mounted on Vertical Motorised Track Slider

Although, a typical barbell squat repetition takes much less time a compromise was made to ensure the acquisition of a sufficient amount of input frames. The configuration of the IR LEDS during the data acquisition process was the same as that in the static case.

5) *Mobile Case – Results*

The results of the experimental validation in the mobile case are shown in TABLE IV. These results correspond to the results of simulations in scenario 3, shown

TABLE *I*. Likewise, a visual representation of the results of this experiment is shown in Fig. 13, which corresponds to results of simulated scenario 3 shown in Fig. 10. These results bear a strong resemblance to those of the corresponding simulations. The RMSE in P_z^W was the highest of the three position elements of the WP. Also, it was higher than that in P_y^W by a comparable ratio of approximately 50 %. Likewise, the RMSE in P_x^W had the lowest value of the three position elements of the pose vector P_{WP}^W.

TABLE IV
MOBILE CASE - EXPERIMENTAL RESULTS

	$P_x^W[mm]$	$P_y^W[mm]$	$P_z^W[mm]$	Total [mm]
RMSE	2.5	11.5	20.4	13.6

XXXX-XXXX © XXXX IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

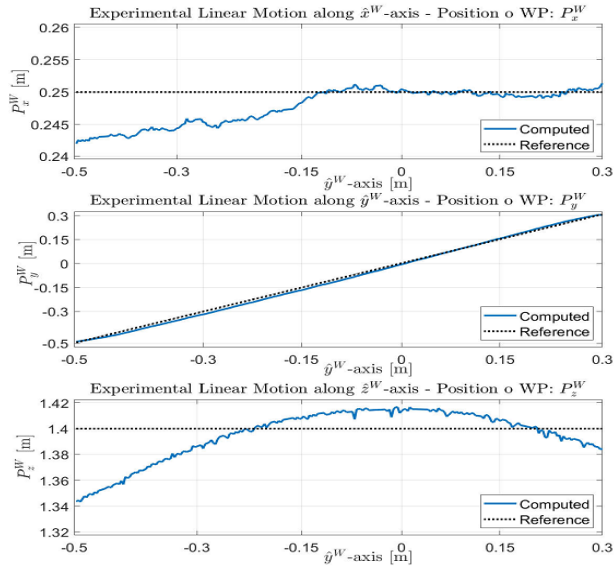

Fig. 13. Experimentally Determined Position of the WP in Linear Motion along $\hat{y}^W$-axis

Overall, the RMSE was lower than that in the corresponding simulated scenario 3. The discrepancy between these results was low and in the order of several millimeters, i.e. less than 5 mm. One of the reasons for such as low value of RMSE is the relatively low velocity of the WP whose period was $T = 34\ s$. Also, the motor controller ensured a smooth change of the motion's direction at the inflection points, i.e. when $y^W = -0.5$ or $y^W = 0.3$. It may have, to some extent, reduced the error in IMU readings. Moreover, this motion pattern involved no rotations, thus making the IMU readings less susceptible to error.

An additional experiment was carried out to determine the repeatability of the proposed system and its algorithm. To this end, the motor controller program on the Raspberry Pi was programmed to drive the WP to perform ten full cycles of scenario 3; to simulate ten repetitions of the barbell squat. Fig. 14 shows the results of this experiment. These results show that the performance of the proposed system was consistent and repeatable in all ten cycles. It is also evident that the output of the IMU did not drift, thus avoiding the adverse impact on the sensor fusion algorithm's accuracy.

IV. Result Analysis And Benchmark Comparison

The proposed system was extensively evaluated both in terms of performance modelling and laboratory experiments. The experimental results confirmed the expectations based on the simulations.

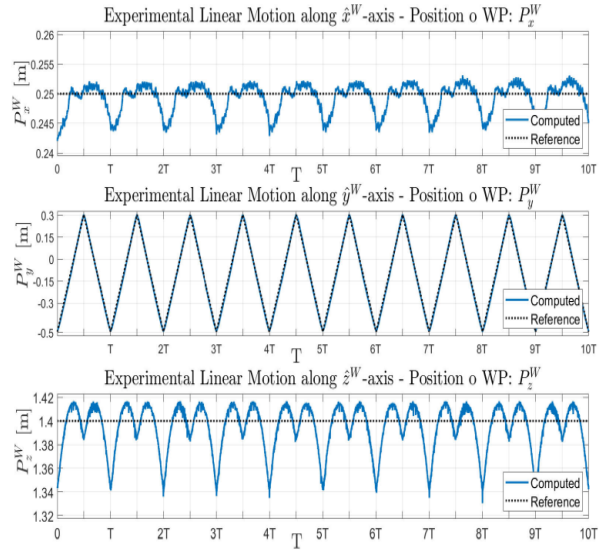

Fig. 14. Experimentally Determined Position of the WP in Linear Motion along $\hat{y}^W$-axis Over Ten Repetitions with T = 34 s

Whereas the predicted RMSE in Scenario 2 and Scenario 3 was 34.5 mm and 19.8 mm, for noise levels $\mathcal{N}_3$ and $\mathcal{N}_2$, respectively. The corresponding experimental results were 36.7 mm and 13.6 mm. These results confirm the expected performance when the system is affected by noise, mainly that originating from the camera and IMU. Moreover, the results of our recent study where we implemented the proposed system on an embedded platform were largely consistent with the mobile case, albeit less accurate due to differences in the between the two systems [41]. The RMSE in 3-D position calculation in the embedded platform was 32.8 mm.

The performance of the proposed novel system described in this paper was compared to similar systems that exist in the SOA for motion tracking. One of the key selection criteria for this comparison was the similarity in terms of system architecture, in particular the use of monocular vision and IMU sensor fusion for pose estimation. A direct one-to-one comparison was not possible due to different performance validation metrics, target application spaces, system architectures, and the algorithms used in these approaches. However, a general comparison can be made between the new system developed and the SOA. TABLE V compares and contrasts some of the key properties of the proposed system to the three most comparable alternatives in the SOA, as reported in the respective referenced publications.

The key metric for comparison was the overall error in position estimation, as well as the parameters that describe the key requirements of the individual systems. The IS-1500 tracker was the most accurate inside-out tracker,

XXXX-XXXX © XXXX IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

whose position error was, by far, the lowest. However, it required at least four fiducial markers and high external computing power capabilities to achieve such results, thus being the most expensive and complex system in this comparison. The motion tracking system proposed by Maereg et al. reported a very low RMSE. However, it achieved such an accuracy within the smallest work envelope of only several centimetres and only in static case, at a single position; while the accuracy in mobile case was not assessed quantitatively. On the other hand, the system proposed by Li et al. had a similar performance to the system proposed in this work. It was also validated in a somewhat similar way to the novel system developed and described in this paper.

The RMSE was determined at a number of static positions along a straight line parallel to the $\hat{z}^W$ axis at distances between 113 and 413 mm. However, the proposed system achieved lower overall RMSE in position estimation in both static and mobile cases, as shown in

TABLE III and TABLE IV, respectively. Both systems had a low complexity. However, the tracker proposed by Li et al. was an outside-in tracker, while our proposed system was and inside-out tracker, which had practical implications related to ease of use and scalability of the working area.

V. Discussion

This study aimed at designing, developing, and evaluating a novel motion tracking system that can be implemented as a small-form-factor wearable device and used resource-constrained wearable applications. It can offer a balanced alternative to the existing alternatives in the literature. The proposed novel system advances the SOA in the following ways. It combines the advantages of the comparable alternatives in the SOA. Firstly, it is an inside-out tracking system. The advantage of an inside-out tracker over the outside-in trackers is in that the size of the work envelope can be scaled at little to no expense. The costliest component, both in terms of price and complexity, is the camera. The proposed system, like the IS-1500, has one monocular camera embedded in the WP, regardless of the size of the work envelope. The algorithm does not change, as long as two points of reference are in camera's FoV and their baseline B is known. Whereas the outside-in systems would require additional cameras to scale the work envelope, the proposed system would need only additional IR LEDs, whose implications in terms of complexity and cost are significantly lower.

Secondly, the proposed system is less complex in terms of the architecture and algorithm, as compared to the IS-1500. In this regard, it is more comparable to the two outside-in alternatives that also rely on two tracking points of reference. In summary, the proposed system has the advantage of the inside-out systems while being less complex and, thus more suitable for low-power, embedded wearable motion tracking devices for various application spaces, such as the barbell squat in ST, for instance.

The main limitation of the proposed system is in that it achieves the highest accuracy when the WP is near the centre of the work envelope and rotates mainly about a single axis while the rotations about the remaining two axes are relatively small. The extreme poses of the WP increase the RMSE, which was shown in the static case of the experimental validation. Nevertheless, the proposed sensor fusion algorithm can handle multi-axis rotations with rotation angles up to approximately $10\,deg$ about each axis, which was shown in the simulated uniform random scenario 2. The proposed algorithm is susceptible to noise in IMU readings. The point noise also affects the performance but to a lesser degree. The impact of noise $\mathcal{N}$ is particularly high in scenarios that involve significant multi-axis rotations, such as that in the simulated scenario 2, whose impact is shown in Fig. 8. Moreover, the RMSE in 3-D position calculation in the static case was higher than that in the mobile case. It was an expected result. Whereas the WP's pose changed in a repeated vertical motion pattern in the mobile case, the static case exercised a number of poses across the entire work envelope; including the extreme positions and orientations.

TABLE V
COMPARISON OF THE PROPOSED SYSTEM TO ALTERNATIVE SOLUTIONS IN THE SOA

	Position Error [mm]	Markers Required	Tracking Type	Work Envelope Size (along z-axis) [m]	Overall System Complexity
IS-1500 (PRA algorithm with Fiducial Markers) [13]	2 (Typical)	At least 4 (Passive Fiducial)	Inside-Out	Variable	High
Maereg et al. [14]	0.21 (Static) (RMSE)	2 (Active)	Outside-in	0.045	Low
Li et al. [15]	48.3 to 275.4 (Static) (RMSE)	2 (Passive)	Outside-In	1.13 to 4.13	Low
Proposed System	36.7 (Static), 13.6 (Mobile) (RMSE)	2 (Active)	Inside-Out	0.5 to 1.5	Low

XXXX-XXXX © XXXX IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

For instance, the poses of the WP that were close to the limits of, i.e. edges of the work envelope and rotations close to 10 degrees about any axis, were expected to have high RMSE, which in turn increased overall RMSE in the static scenario.

VI. Conclusions

In this work, a novel prototype system for low-power, miniaturized, wearable human motion tracking devices for sports applications was presented. The novelty of the proposed system is twofold. The first novelty is in the system architecture. Namely, the proposed system is an inside-out motion tracker that comprises of the WP with two complementary sensor modalities, i.e. a monocular camera and an IMU sensor, and two points of reference embedded in the ambient environment, i.e. IR LEDs. Furthermore, the proposed multimodal sensor fusion algorithm is novel in how it computes the 3-D pose using information from the two sensor modalities along with only two external reference points. The sensor fusion algorithm is executed on the WP, and it leverages the complementary nature of the monocular vision and IMU sensor modalities to directly compute the 3-D pose of the WP. The target application spaces for this system include sports applications. It can be particularly applicable to tracking certain exercises in ST routines, such as the barbell squat.

The proposed system was implemented and validated in the form of a prototype experimental setup in laboratory conditions. Its performance was experimentally validated in two scenarios. The static case was aimed at determining the performance across the entire work envelope. The major findings in this work include the accuracy of the system given its relative simplicity as compared to other comparable alternatives in the SOA. The RMSE in 3-D position calculation was 36.7 mm and 13.6 mm in the static and mobile cases, respectively. The mobile case focused on the motion pattern that is normally involved in a barbell squat. This scenario was of main interest, as this system is intended to be used in tracking such motion patterns when it has moved to the next development stage, i.e. a small-form-factor prototype stage implementation.

The proposed system did not match the performance of the IS-1500 inside-out tracker in terms of accuracy. The IS-1500 had the lowest error of all comparable systems present in the SOA. However, the IS-1500 is a complex system with high processing requirements. For example, it requires at least four reference points (fiducial markers) and an externally connected processing power capability, e.g. a laptop PC. On the other hand, the proposed system compared well to the other two outside-in tracking systems, as shown in TABLE V. It needs to be noted, though, that the RMSE of these two systems cannot be directly compared due to different validation scenarios. Therefore, the experimental conditions need to be also taken into account. Nevertheless, the proposed system performed better than that proposed by Li et al. Although the monocular version of their system was validated at a set of static positions along a straight horizontal line with no rotations, RMSE of the system proposed was lower in both experimental scenarios. On the other hand, the outside-in tracker proposed by Maereg et al. achieved lower RMSE. However, it achieved this result within a much smaller work envelope in static conditions with no rotations.

The viability of the proposed system in the context of low-power embedded motion tracking applications was evaluated in our recent study [41]. The system architecture and multimodal sensor fusion algorithm were implemented using a low-cost embedded platform. The RMSE and execution time were determined using a mobile scenario that was similar to that described in this work, albeit not identical. The total RMSE in position calculation was 32.8 mm. The embedded WP operated in real-time at just over 20 FPS. It needs to be noted that, the RMSE in position tracking of the embedded WP was not expected to match that of the corresponding mobile scenario described in this work; due to different properties of the experimental setups, such as the hardware properties of the WP or work envelope.

The results of this work show that the proposed motion tracker may be an alternative to human motion tracking using wearable devices that is worth considering; especially in low-cost applications. It is an inside-out opto-inertial motion tracker that performs 3-D pose detection using only two points of reference in the ambient environment. It is a less expensive, simpler, and more scalable approach, as compared to the alternatives present in the SOA, such as the IS-1500. On the other hand, the two outside-in trackers considered in this work are less scalable, while being similar conceptually. Also, their usability in the context of wearables is limited by the fact that their accuracy is also affected by the distance between the two points of reference, which must be small if these were to be attached to the human body. Moreover, small distance between the reference points, in conjunction with considerable distance away from the camera, increases the cost of the system, due the requirement of a higher camera resolution to compensate for that. Thus, the proposed tracker advances the SOA by proposing a balanced alternative to the existing systems, albeit not as accurate as the leading IS-1500. However, it can be considered a viable alternative if other factors are taken into account, such as the scalability. Moreover, the proposed system achieved a sufficiently low error in position estimation to be good enough for tracking human motion in certain exercises, such as the barbell squats in ST routines. The proposed system can be used in real-world cases as follows. The outputs of the multiple WPs, distributed on the athlete's back, can be used as inputs to a higher-level system, also a wearable one. The high-level system could assess the correctness of the exercise execution based on the motion patterns; represented by the sensor data form the WPs. There can be an element of machine learning and edge computing involved in this part. The athlete and receive the feedback in near-real-time via AR headset.

Future work can involve the development of a small-form-factor proof-of-concept prototype of the WP which will turn this pre-prototype system into a miniature wearable device. Furthermore, additional experimental evaluation of the miniaturized WP can be carried out with the involvement of human subjects. The individuals can perform real barbell squats with the miniaturized WP attached to their back.

VII. References

[1] S. O. H. Madgwick, A. J. L. Harrison, and R. Vaidyanathan, "Estimation of IMU and MARG orientation using a gradient descent algorithm," in

2011 IEEE International Conference on Rehabilitation Robotics, June 29 2011-July 1 2011 2011, pp. 1-7, doi: 10.1109/ICORR.2011.5975346.

[2] L. Naimark, E. Foxlin, and I. C. Soc, "Encoded LED system for optical trackers," *International Symposium on Mixed and Augmented Reality, Proceedings*, pp. 150-153, 2005.

[3] R. States and E. Pappas, "Precision and repeatability of the Optotrak 3020 motion measurement system," *Journal of medical engineering & technology*, vol. 30, no. 1, pp. 11-16, 2006.

[4] M. Windolf, N. Götzen, and M. Morlock, "Systematic accuracy and precision analysis of video motion capturing systems—exemplified on the Vicon-460 system," *Journal of biomechanics*, vol. 41, no. 12, pp. 2776-2780, 2008.

[5] J. A. Faulkner, L. M. Larkin, D. R. Claflin, and S. V. Brooks, "AGE-RELATED CHANGES IN THE STRUCTURE AND FUNCTION OF SKELETAL MUSCLES," *Clinical and Experimental Pharmacology and Physiology*, vol. 34, no. 11, pp. 1091-1096, 2007/11/01 2007, doi: 10.1111/j.1440-1681.2007.04752.x.

[6] P. Comfort and P. Kasim, "Optimizing squat technique," *Strength and Conditioning Journal*, vol. 29, no. 6, p. 10, 2007.

[7] C. Bishop and A. Turner, "Integrated Approach to Correcting the High-Bar Back Squat From “Excessive Forward Leaning”," *Strength & Conditioning Journal*, vol. 39, no. 6, pp. 46-53, 2017.

[8] R. F. Escamilla *et al.*, "Effects of technique variations on knee biomechanics during the squat and leg press," *Medicine & Science in Sports & Exercise*, vol. 33, no. 9, pp. 1552-1566, 2001.

[9] T. M. McLaughlin, T. J. Lardner, and C. J. Dillman, "Kinetics of the parallel squat," *Research Quarterly. American Alliance for Health, Physical Education and Recreation*, vol. 49, no. 2, pp. 175-189, 1978.

[10] G. D. Myer *et al.*, "The back squat: A proposed assessment of functional deficits and technical factors that limit performance," *Strength and conditioning journal*, vol. 36, no. 6, p. 4, 2014.

[11] K. Peterson, "Muscle Activation and Range of Motion Patterns of Individuals who Display a Lateral Hip Shift during an Overhead Squat," Masters Thesis, University of North Carolina at Chapel Hill Graduate School, 2019.

[12] D. Diggin *et al.*, "A biomechanical analysis of front versus back squat: injury implications," in *ISBS-Conference Proceedings Archive*, 2011, vol. 1, no. 1.

[13] Intersense. "IS-1500." http://www.intersense.com/pages/70/255 (accessed 25.03.2019, 2019).

[14] A. Maereg, E. Secco, T. Agidew, D. Reid, and A. Nagar, "A Low-Cost, Wearable Opto-Inertial 6-DOF Hand Pose Tracking System for VR," *Technologies*, vol. 5, no. 3, p. 49, 2017.

[15] J. Li, J. A. Besada, A. M. Bernardos, P. Tarrío, and J. R. Casar, "A novel system for object pose estimation using fused vision and inertial data," *Information Fusion*, vol. 33, pp. 15-28, 2017/01/01/ 2017, doi: http://dx.doi.org/10.1016/j.inffus.2016.04.006.

[16] X. Xu and R. W. McGorry, "The validity of the first and second generation Microsoft Kinect™ for identifying joint center locations during static postures," *Applied ergonomics*, vol. 49, pp. 47-54, 2015.

[17] A. Pfister, A. M. West, S. Bronner, and J. A. Noah, "Comparative abilities of Microsoft Kinect and Vicon 3D motion capture for gait analysis," *Journal of Medical Engineering & Technology*, vol. 38, no. 5, pp. 274-280, Jul 2014, doi: 10.3109/03091902.2014.909540.

[18] D. R. P. S. H. Luinge, "Xsens MVN: Full 6DOF Human Motion Tracking Using Miniature Inertial Sensors," XCENS Technologies, 2013.

[19] B. O'Flynn *et al.*, "Novel smart glove technology as a biomechanical monitoring tool," *Sensors & Transducers*, vol. 193, no. 10, p. 23, 2015.

[20] G. Welch and E. Foxlin, "Motion tracking: No silver bullet, but a respectable arsenal," *IEEE Computer graphics and Applications*, vol. 22, no. 6, pp. 24-38, 2002.

[21] E. Foxlin and L. Naimark, "VIS-Tracker: A wearable vision-inertial self-tracker," in *IEEE Virtual Reality 2003 Conference*, Los Angeles, Ca, Mar 22-26 2003, in Proceedings of the IEEE Virtual Reality Annual International Symposium, 2003, pp. 199-206, doi: 10.1109/vr.2003.1191139. [Online]. Available: <Go to ISI>://WOS:000182252300026

[22] E. Foxlin, T. Calloway, and H. Zhang, "Improved registration for vehicular AR using auto-harmonization," in *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 10-12 Sept. 2014 2014, pp. 105-112, doi: 10.1109/ISMAR.2014.6948415.

[23] R. Atac, S. Spink, T. Calloway, and E. Foxlin, "Scorpion Hybrid Optical based Inertial Tracker (HObIT) Test Results," in *Conference on Display Technologies and Applications for Defense, Security, and Avionics VIII*, Baltimore, MA, May 07-08 2014, vol. 9086, in Proceedings of SPIE, 2014, doi: 10.1117/12.2050363. [Online]. Available: <Go to ISI>://WOS:000343864900018

[24] P. Gill and T. Vogelsang, "Lensless Smart Sensors: Optical and thermal sensing for the Internet of Things," in *2016 IEEE Symposium on VLSI Circuits (VLSI-Circuits)*, 15-17 June 2016 2016, pp. 1-2, doi: 10.1109/VLSIC.2016.7573486.

[25] S. J. Byrnes, A. Lenef, F. Aieta, and F. Capasso, "Designing large, high-efficiency, high-numerical-aperture, transmissive meta-lenses for visible light," *Optics Express*, vol. 24, no. 5, pp. 5110-5124, Mar 2016, doi: 10.1364/oe.24.005110.

[26] M. P. Wilk, A. Urru, S. Tedesco, and B. O. Flynn, "Sub-pixel point detection algorithm for point tracking with low-power wearable camera systems: A simplified linear interpolation," in *2017 28th Irish Signals and Systems Conference (ISSC)*, 20-21 June 2017 2017, pp. 1-6, doi: 10.1109/ISSC.2017.7983629.

[27] L. Abraham, A. Urru, M. P. Wilk, S. Tedesco, M. Walsh, and B. O'Flynn, "Point tracking with lensless smart sensors," in *SENSORS, 2017 IEEE*, 2017: IEEE, pp. 1-3.

[28] B. Siciliano and O. Khatib, *Springer handbook of robotics*. Springer, 2016.

[29] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22, 2000.

[30] Z. Taylor. "Project 3D into 2D image coordinates using a camera model." MathWorks. https://uk.mathworks.com/matlabcentral/fileexchange/48752-project-3d-into-2d-image-coordinates-using-a-camera-model?s_tid=prof_contriblnk (accessed 31.08.2019, 2019).

[31] T. InvenSense, "MPU-9250, Nine-Axis (Gyro+ Accelerometer+ Compass) MEMS MotionTracking™ Device," ed, 2014.

[32] E. O. Inc. "1" Diameter, Optical Cast Plastic IR Longpass Filter." Edmund Optics Inc. https://www.edmundoptics.com/optics/optical-filters/longpass-edge-filters/1quot-diameter-optical-cast-plastic-ir-longpass-filter/ (accessed 19.02.2018, 2018).

[33] I. Vishay Intertechnology. "High Speed Infrared Emitting Diode, 940 nm, VSMB11940X01." http://www.farnell.com/datasheets/2245170.pdf?_ga=2.184115750.1623767235.1519051613-1677736993.1519051613&_gac=1.15708484.1519051613.Cj0KCQiAiKrUBRD6ARIsADS2OLlYcKGvXIt139L6uYIypwMFTqdQy6r-30rp9fSntK9OXyQTsLnW8b8aAm_VEALw_wcB (accessed 19.02.2018, 2018).

[34] Manfrotto. "Manfrotto MN755XB Aluminium Video Tripod." https://www.barkerphotographic.ie/manfrotto-mn755xb-tripod-legs/camera-accessories/tripods (accessed 25.02.2019, 2019).

[35] Manfrotto. "Manfrotto 410 Junior Geared head." https://www.barkerphotographic.ie/manfrotto-410-junior-geared-head/camera-accessories (accessed 25.02.2019, 2019).

[36] "FIXKIT Mini Digital Inclinometer Angle Finder." Online. https://picclick.fr/FIXKIT-Mini-Digital-Inclinometer-Angle-Finder-Gauge-Level-142643236661.html (accessed 25.02.2019, 2019).

[37] "CPTCAM CP-80S Laser Distance Meter." https://es.tmart.com/CPTCAM-CP-80S-Portable-Handheld-80m-Mini-Laser-Rangefinder-Distance-Measuring-Meter-Black-Yellow_p338374.html (accessed 25.02.2019, 2019).

[38] Neewer. "Neewer 80 cm Aluminium Camera Track Slider Rail." https://www.amazon.co.uk/gp/product/B06Y46H989/ref=ppx_yo_dt_b_asin_title_o02_s02?ie=UTF8&psc=1 (accessed 9.08.2019, 2019).

[39] "NEMA 17 - 17HS4401 Stepper Motor." https://datasheet4u.com/datasheet-pdf-file/928661/MotionKing/17HS4401/1 (accessed 9.08.2019, 2019).

[40] "TB6600 Stepper Motor Driver." http://olimex.cl/website_MCI/static/documents/TB6600_data_sheet.pdf (accessed 9.08.2019, 2019).

[41] M. P. Wilk, M. Walsh, and B. O'Flynn, "Low cost embedded multimodal opto-inertial human motion tracking system," in *2020 31st Irish Signals and Systems Conference (ISSC)*, 2020: IEEE, pp. 1-6.

Dr Mariusz P. Wilk earned his BEng (Hons) degree in Electronic Engineering form Cork Institute of Technology, Cork, Ireland in 2013. He earned his PhD degree with the Wireless Sensor Networks Group in Tyndall National Institute, University College Cork, Ireland in 2020. His area of research is in the algorithms for low power wearable technologies. He focuses primarily on multi-sensor data fusion for wearable human body tracking applications that can be used with the emerging state-of-the-art inertial and visual sensor technologies. He currently holds a role of a CTO in a technology start-up company called Watertools where he applies his expertise to solving the global challenge of water scarcity.

Dr Michael Walsh is a Senior Researcher in Wireless Sensor Networks Groups at the Tyndall National Institute. He is also a Co-Founder and CTO at Grasp Wearable Technologies which is a start-up enterprise developing breakthrough IP technologies, applications, and product concepts, targeting international markets in fitness, sports, health, and gaming. His research interests include Wearable technologies for healthcare and sports science, High Performance Wireless Inertial Measurement, Wireless and Mobile Healthcare, Ambient Assisted Living, Next Generation Wireless Technologies, Indoor and Outdoor Real-time Localization Systems.

Dr Brendan O'Flynn is a Senior Staff Researcher at the Tyndall National Institute and is Head of Group for Wireless Sensor Networks Group. Since 2004 he has been defining and leading the research activities of the Wireless Sensor Networks (WSN) group. This multidisciplinary group consists of over 20 researchers comprising of postdoctoral researchers, postgraduate students, engineers, and scientists. He is a funded investigator at CONNECT investigating the Internet-of-Things. His current research interests include System integration/miniaturized sensing systems, Hardware/software co-design of embedded systems and the Deployment of reliable and robust complex systems.